INSTANTANEOUS FREQUENCY ESTIMATION OF SPEECH RESONANCES

A Thesis Submitted

in Partial Fulfilment of the Requirements

for the Degree of

MASTER OF TECHNOLOGY

by

N.V.CHALAPATHI RAO

to the

DEPARTMENT OF ELECTRICAL ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY KANPUR

March, 1998

CERTIFICATE

This is to certify that the work "Instantaneous Frequency Estimation of Speech Resonances", has been carried out by N.V.Chalapathi Rao under my supervision, and it has not been submitted elsewhere for a degree.

Preeti Rao

Dr. Preeti Rao
Asst. Professor
Dept. of Electrical Engg.
Indian Institute of Technology
Kanpur

ACKNOWLEDGMENTS

I express my respect and gratitude to my supervisor Dr. Preeti Rao for her active guidance at every stage of thesis work, it is only with her constant encouragement the present work was completed.

I am, forever, indebted to my parents for their blessings and encouragement, which has helped me to complete the present work smoothly.

I wish to thank all those with whom I had a nice time at IIT kanpur.

2nd March,1998 N.V.Chalapathi Rao

kanpur

ABSTRACT

In the production of voiced speech the glottal source and the vocal tract interact, giving rise to variations in the formant frequencies and bandwidths over the duration of a single pitch period. Spectral correlation techniques cannot be applied to analyze these variations, as they are limited by the available number of data samples. In an attempt to overcome these limitations, a new approach is tried in this work to track these variations of formant frequencies using an instantaneous frequency estimation technique based on the analytic signal method. The estimate is unreliable near the glottal closure and opening instants while there is a clear increase in the frequencies of the first and second formants in the glottal open phase of some cases.

Contents

List of Figures

Chapter 1

Introduction

1.1 Development of DSP Systems and Algorithms

Before 1960 Digital Signal Processing (DSP) was mainly done with the help of processors which were hardware-built and task-devoted and had long computation times. This limited their further development in future applications. With the advent of IC's in mid and late 1960's digital signal processing picked up speed as the design of processors specific to signal processing tasks became possible. In the field of IC's we now see very large scale, very high speed IC's which do long computations in short times. This led to the development of general signal processing algorithms which perform the required tasks quickly and efficiently. The computer simulations became easy with the onset of PC revolution and helped further in the development of these algorithms.

1.2 Role of Speech Analysis in Speech Processing Systems

Speech signal processing forms one of the main applications of digital signal processing in addition to other fields like image processing, robotics, seismic signal processing, mobile communication systems, etc. It is mainly comprised of applications in speech coding, speech synthesis, speech recognition and speaker recognition. The speech coding system when coupled to a corresponding speech synthesis system gives rise to a vocoder, which can provide the efficient transmission or storage of speech signals. The speech recognition systems are used to recognize the text spoken by the speaker with out worrying about the person who has spoken it. The speaker recognition system is useful in recognizing the speaker without considering the text he has spoken. The speech analysis plays a major role as a front end for all these systems and determines the systems efficiency in its applications. In speech coding systems, speech analysis research aims at reducing the output bit-rate by obtaining compact parameterizations of the speech signal, preserving the speech quality. In speaker recognition systems it helps in computing the speaker identity feature vectors which are used in recognising the speaker. The speech analysis is also utilised in many aids to the handicapped like the design of sensory aids and visual displays of speech information, which are used in teaching deaf persons to speak. It also is helpful in enhancing the quality of speech signals degraded by noise.

1.3 Brief Description of Problem

The ability to automatically find and track the formant frequencies(resonance frequencies of the vocal tract tube) is an important part of speech processing systems because formants play a major role in most speech analysis applications. The traditional methods for formant finding are peak picking of the cepstrally-smoothed or

LPC Spectrum, or finding the roots of the LPC polynomial. These methods assume that the formants are constant with in an analysis frame. The speech synthesized using a model which assumes constant formant frequency in a frame, is generally very intelligible but often sounds unnatural. The formant frequencies can be taken to be constant if the assumption that the glottal source and the vocal tract are linearly seperable is true. But it is found that they actually interact and this leads to variation of formant frequency even in a single pitch period. Hence knowledge of the variation of the formant frequency in a single pitch period is necessary to develop models which can solve the problem of unnaturalness in the synthesized speech. We made an attempt here to understand the formant frequency variations of the speech signal in a single pitch period using the analytical signal method of determining the instantaneous frequency of a narrow band signal.

1.4 Organization of the Chapters

In Chapter 2 the review of the literature study is presented which motivated us to take up the present problem. Chapter 3 gives the description of analytic signal method and it's advantages over other methods for calculating the instantaneous frequency. A brief description about the implemented algorithm and the results, observations and conclusions as applied to speech signal are presented in Chapter 4. A package developed to carry out the procedure given in chapter 4 is described in Appendix.

Chapter 2

Motivation for taking up the Present Problem

2.1 Speech Production Model

We know that the speech signal is simply the electrical equivalent of the acoustic wave that is radiated from mouth when air is expelled from the lungs and the resulting flow of air is perturbed by a constriction somewhere in the vocal tract. It is found that a parametric representation of the speech signal (i.e. representing the speech signal as the output of a model of speech production) is advantageous in all the applications of speech analysis, compared to waveform representation which simply preserves the waveshape of the analog speech signal. The basic model that is typically used for the speech production mechanism is shown in the Figure 2.1.

The basic speech production model is based on the observation that speech sounds can be classified into two distinct classes according to their mode of excitation. The parameters of this model are conveniently classified as either excitation parameters(i.e. related to the source of speech sounds) or vocal tract response pa-

rameters (i.e. related to individual speech sounds). Speech analysis is the process of estimating the (time varying) parameters of the model for speech production from a speech signal that is assumed to be the output of that model. It also plays a major role in improving the existing models in terms of speech quality (by improving naturalness and intelligibility) and efficiency(reducing the output bit-rate). If the model is sufficiently accurate and the parameters are accurately determined, the resulting output of the model is in some cases indistinguishable from natural speech.

The voiced sounds are produced by forcing air through the glottis with the tension of the vocal cords adjusted so that they vibrate in a relaxation oscillation, there by producing quasi-periodic pulses of air which excite the vocal tract. The quasi-periodic waveform which is the output of glottal source is termed glottal volume-velocity waveform (shown in Figure 2.3) and modelled using glottal pulse parameters. We can observe the glottal closure and the glottal opening times from this figure. As sound propagates down the vocal tract tube, the frequency spectrum is shaped by the frequency selectivity of this tube. The resonance frequencies of this tube are called formant frequencies or simply formants. These formants depend upon the shape and dimensions of the vocal tract; each shape is characterized by a set of formant frequencies. Different sounds are formed by varying the shape of the vocal tract. Thus the spectral properties of the speech signal vary with time as the vocal tract shape varies. The output of the vocal tract forms the lip volume-velocity waveform from which the speech signal (i.e. lip pressure waveform) is obtained after the differentiating action of the lips.

To calculate glottal pulse parameters, an inverse filtering method is generally followed [2] in which the speech signal is first inverse filtered (corresponds to removing the effect of vocal tract system). The resulting signal represents the differentiated glottal volume-velocity waveform and is shown in Figure 2.3. The integration of this waveform yields the required glottal pulse parameters from the obtained volume-velocity waveform.

In the above speech production model it is assumed that the glottal source and the vocal tract do not interact. The formant frequencies are taken to be constant in each pitch period as the change in the dimensions of vocal tract is negligible in this time. However it is found that the glottal source and the vocal tract do interact [15] and this results in the variation of formant frequency even within a single pitch period. Incorporating source-tract interaction into the speech production model is found to lead to an increase in the naturalness of the synthesized speech. Let us briefly review some of the methods which attempted to study this source-tract interaction in the next section.

2.2 Previous Studies on Source-Tract Interaction

It is found that [2] the methods used to simulate the effects of source-tract interaction may be classified as either interactive or non-interactive. The interactive model does not separate the glottal source and the vocal tract. The interaction of these two systems is modeled by a non-linear, time varying model. But the interactive approach for simulating source-tract interaction requires a knowledge of the parameters that are not easily measured and this approach is not readily implemented in a speech synthesizer. The non-interactive approach models the source and vocal tract filter as linearly separable systems with time varying parameters that approximate the source-tract interaction. The speech production model using non-interactive approach is given in Figure 2.2.

In the case of the non-interactive speech production approach two major effects are found necessary to be included. One is the skewing of the glottal volume-veloctity to the right with respect to the glottal area as shown in Figure 2.4. This effect is caused by the vocal tract loading the source The second is the superimposition of the ripples on the opening phase segment of the glottal volume-velocity

waveform. This effect is shown in Figure 2.5 and these ripples have been attributed to the first formant energy which is dissipated by the glottis during the open phase of the glottal cycle. In [2] the authors have used inverse filtering approach to estimate glottal volume-velocity waveform parameters. To simulate the ripple effect they have increased first and second formant bandwidths by a factor of four for the open interval over that of the closed interval.

From the above, we know that even in the steady voiced sounds, excitation characteristics change within each pitch period due to glottal vibrations and the vocal tract system changes due to coupling and decoupling of the trachea during open and closed phases of the glottal excitation. Linear prediction coefficients capture only the averaged behavior over the analysis frame. Accordingly, the detail lost in LPC modeling cannot be easily compensated for by using a glottal pulse for the excitation model. Hence the idea to study more extensively the frequency response behaviour of the vocal tract over a single pitch period came into being. The main difficulty is in determining the characteristics of the vocal tract system from short (2-4 msec) segments of the speech signal, in the two distinct phases in each pitch period, namely, the closed and open glottis regions. In paper [6] the authors developed methods to reduce the effects of the short window in the analysis using a method called source-system windowing. They observed a significant increase in the bandwidth of the first formant and an increase in the value of the first formant in some cases in the open glottis region. Through informal listening, they have noticed that synthesizing speech using separate LPC's for open and closed glottis regions produces a more natural sounding speech compared to conventional LPC synthesis using a glottal pulse model for excitation.

2.3 The Frequency Estimation Approach

Teager [16] found that band-pass filtering the speech vowels around their formants and then applying the energy operator method often yielded several pulses which he called "energy pulses" per pitch period. He reasoned out that these energy pulses indicate some kind of modulation in each formant. This motivated the authors in [3] to model the speech resonances by AM-FM model, which led them to study the vocal tract frequency characteristics by directly trying to track the frequency variations of speech resonances sample by sample. However, this method presents certain disadvantages which led us to choose the analytic signal based instantaneous frequency estimation method to try to track the speech resonances. In order to relate the variation in the frequency of the speech formant with the underlying glottal excitation and its interaction with the vocal tract, we have also estimated the glottal closure and opening instants from the speech waveform. This is done using the approximate glottal closure opening instant algorithm presented in [5].

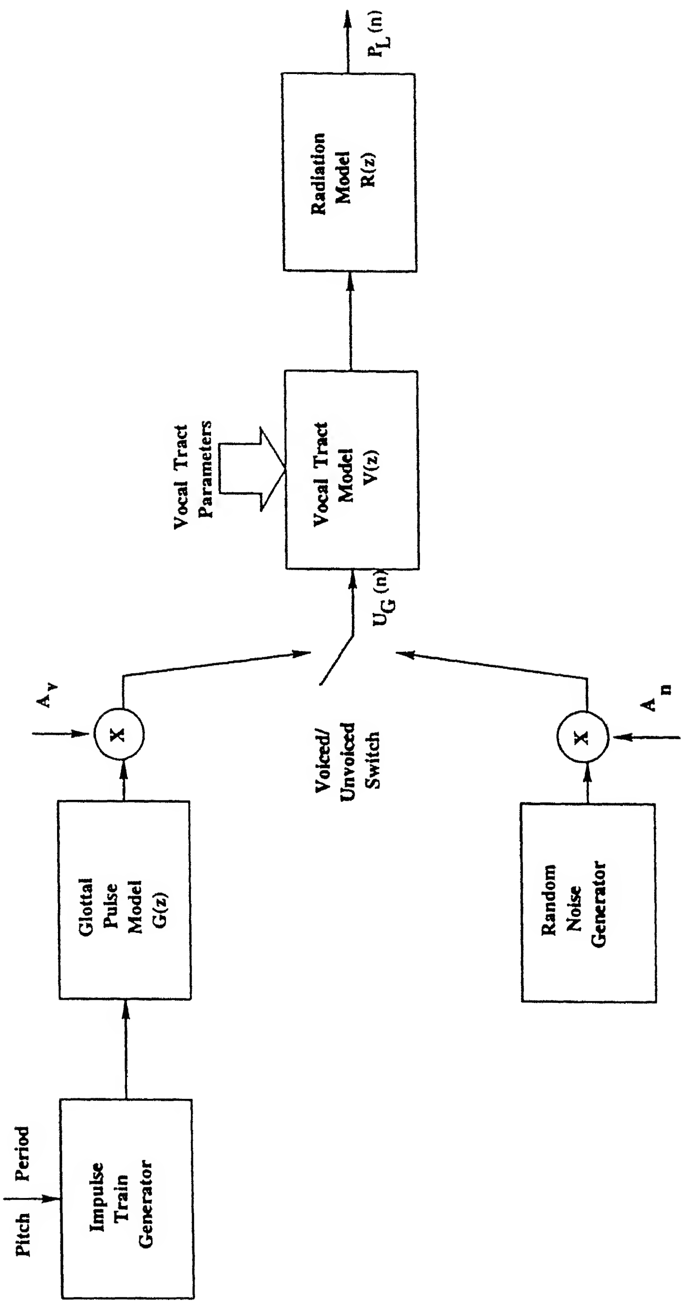

Figure 2.1: Basic Speech Production Model

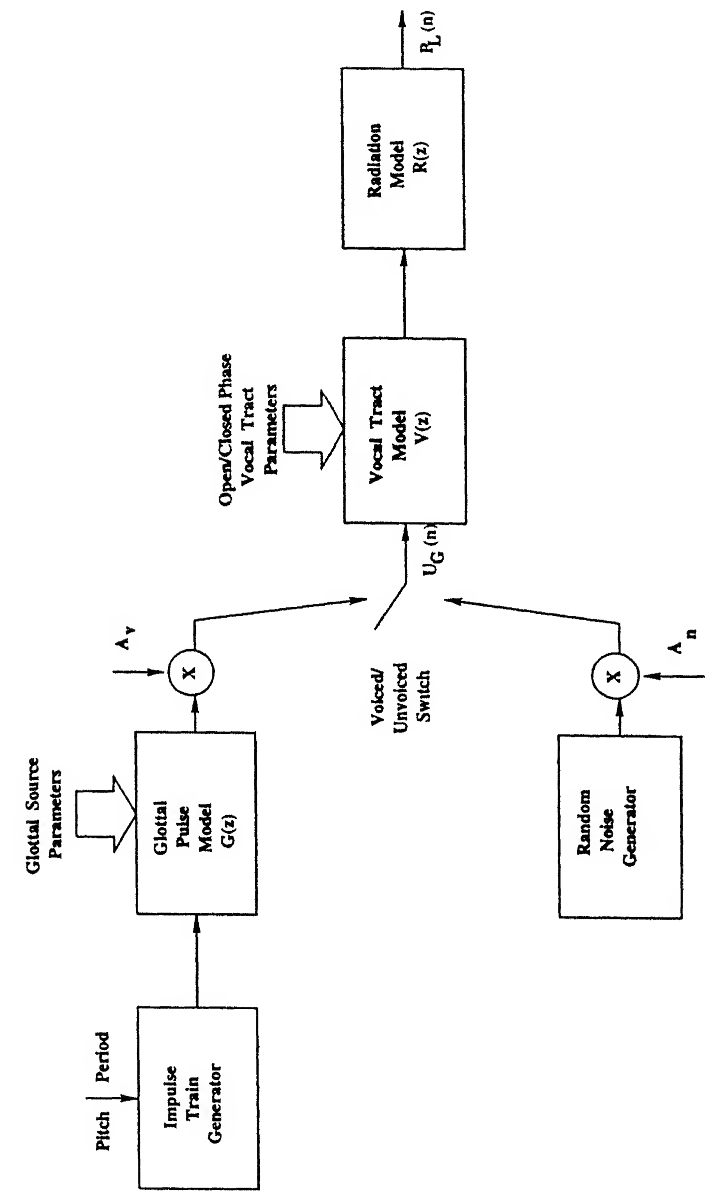

Figure 2.2: Speech Production Model with Source-Tract Interaction

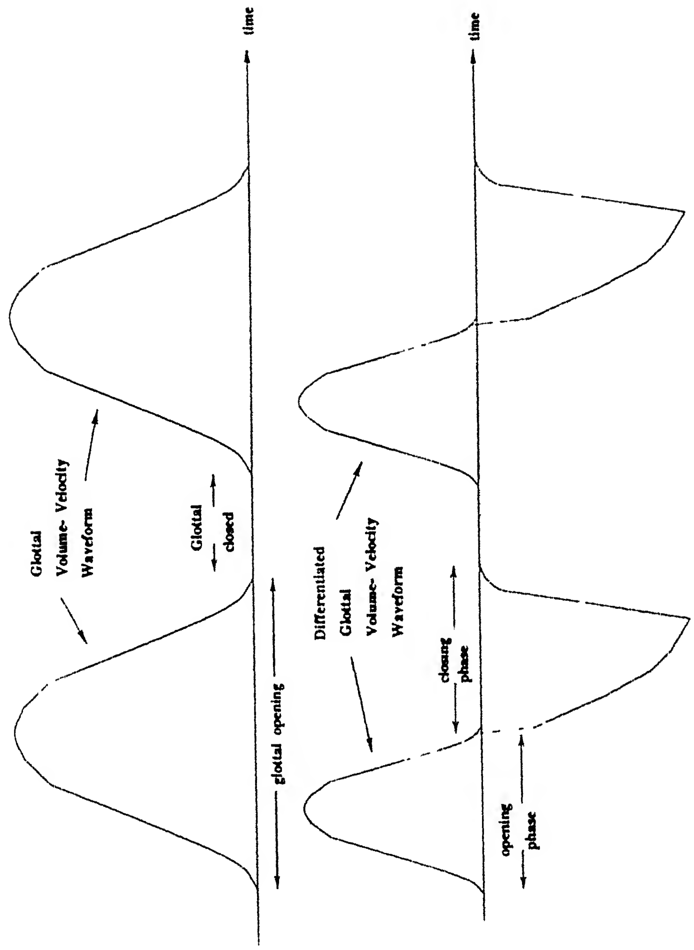

Figure 2.3: Glottal volume-velocity and Differentiated volume-velocity waveforms

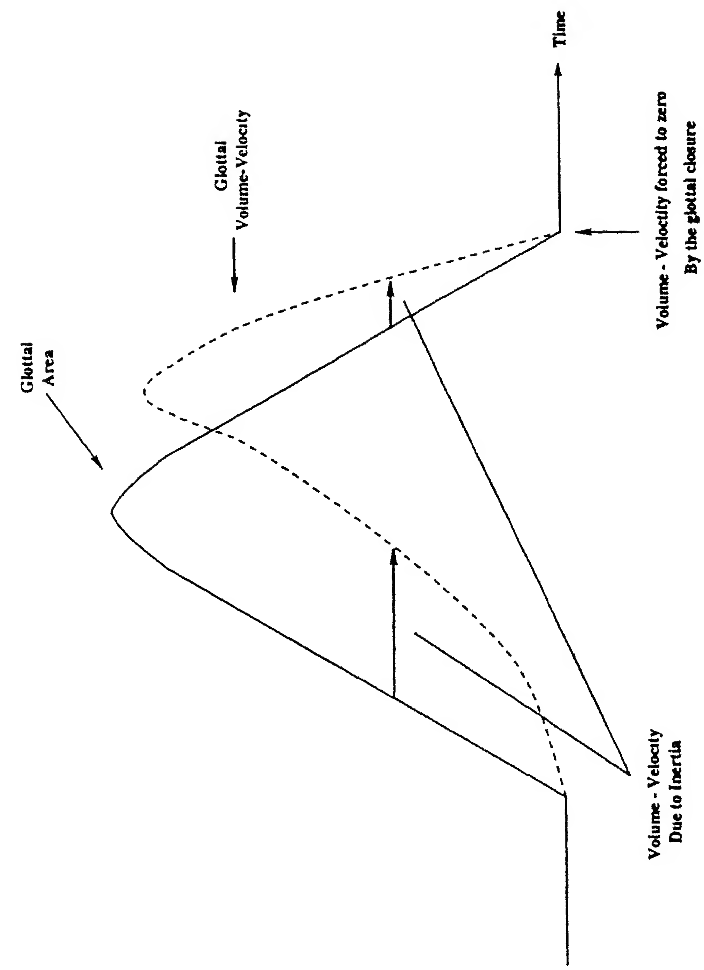

Figure 2.4: Glottal volume-velocity skewed to the right

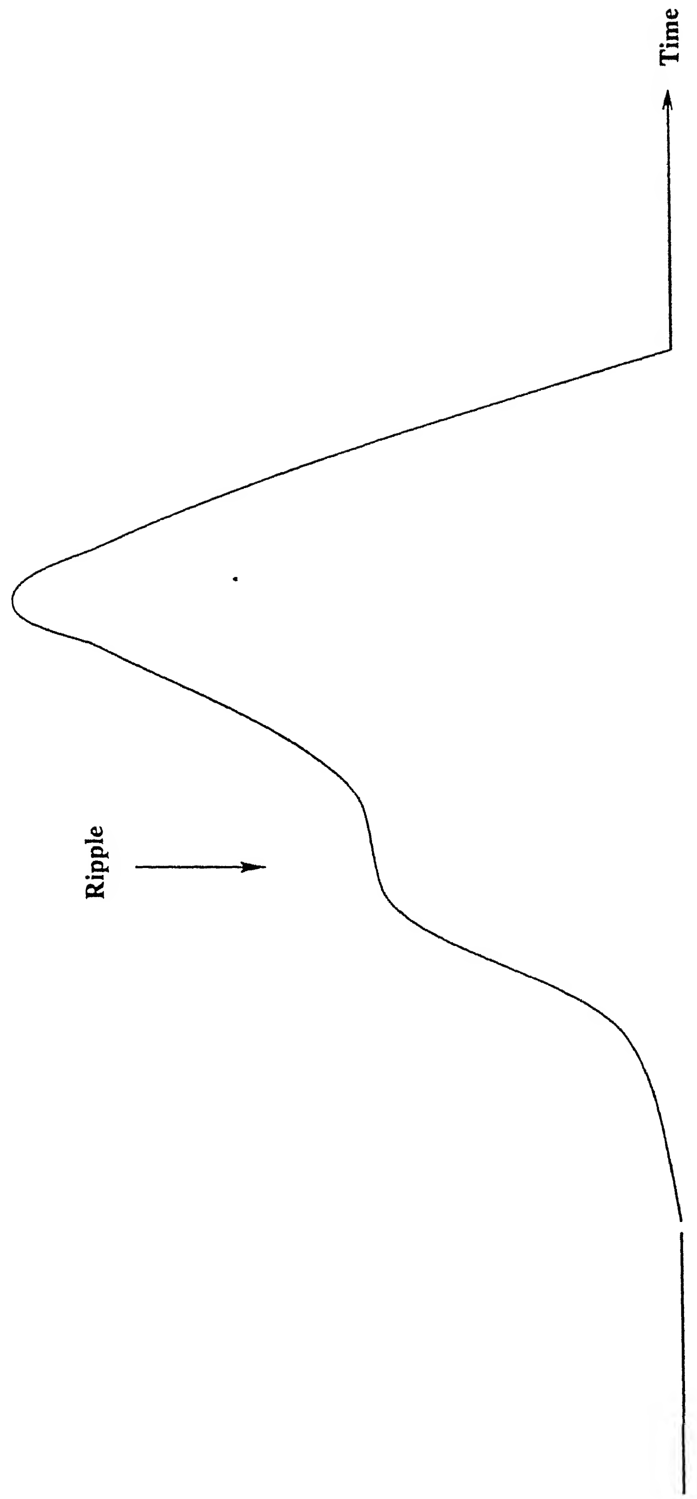

Figure 2.5: Ripple superimposed on the glottal volume-velocity waveform

Chapter 3

Analytic Signal Method of Finding the Instantaneous Frequency of a Signal

3.1 Definition of IF in the Analytic Signal Method

If u(t) is the real signal it is often very advantageous to write complex signal w(t) as

$$w(t)=u(t)+i.v(t)$$

$$=a(t).e^{\phi(t)}$$

is constructed by adding an imaginary part v(t) to the real signal u(t). The amplitude, phase and frequency are defined as

$$a^2(t) = u^2 + v^2$$

$$=| w |^2,$$

$$\phi(t) = arctan[\tfrac{v}{u}]$$

$$=\text{Arg}[w],$$

$$\omega(t) = \tfrac{v'.u - u'.v}{u^2+v^2}$$

$$= Im[\tfrac{w'}{w}].$$

The imaginary part of the signal v(t) is related to real part u(t) by v(t)=H[u(t)] and there are an infinite no of possibilities for H. It can be shown that [1] Hilbert transform operator defined as

$$H[u(t)] = \tfrac{1}{\pi} \int_{-\infty}^{\infty} \tfrac{u(s)}{t-s}.ds$$

is the only valid choice if certain reasonable physical conditions are to be satisfied by the amplitude, phase and frequency.

3.2 Comparison of Analytic Signal Method with Other Methods

To solve the problem of estimating the amplitude envelope and instantaneous frequency of an AM-FM signal the authors in [4] developed an approach that uses an energy operator to separate the signal's output energy product into its amplitude modulation and frequency modulation components. In this paper the authors prefered the energy operator method to the analytic signal method of finding the instantaneous frequency of formant frequency filtered speech signal since the results are almost similar and the computational requirement is less in the former method. They also made the following observations

1. In some cases the analytic signal algorithm seems to yield slightly smoother estimates than the energy operator method.

2. Also in a few isolated instances energy operator method may produce narrow spikes. (eg. at envelope minimum and at the corresponding places of the instantaneous frequency estimates)

The above disadvatages of the energy operator method stand in our way of clearly understanding the formant frequency variations (computational requirement is not a restriction as we are presently interested in understanding the variations of formant frequency which can help us in improving the naturalness of synthetic speech in particular and speech analysis applications in general).

3.3 The Analytic Signal Algorithm

The discription of the analytic signal method algorithm to be applied to the digital narrow band signal is given as below

step 1. The input spectrum V(f) of real signal v[n](sampled version of v(t)) is computed using FFT.

step 2. The analytic signal w(n) of v(n) is calculated by forcing the magnitudes of negative frequencies to zero and doubling the magnitudes of positive frequencies of the calculated FFT in step 1. and taking inverse FFT of the modified spectrum.

step 3. The derivative of the analytic signal w'(n) is computed from the modified spectrum by taking the inverse FFT of product of modified spectrum and iw.

step 4. Instantaneous frequency is calculated using the formula

w(t)=Im[w'(n)/w(n)]

3.4 Simulation Results and Conclusions

The above stated analytical signal method algorithm was implemented using Ansi C programming language on Unix based HP-UX Workstation. To verify the simulation, known AM-FM modulated signals are generated and applied to the program. We can see that the program vak.c works well for constant amplitude, constant frequency sinusoidal signal(case(i) shown in Fig 3.1). Since we are interested in applying this algorithm to the voiced speech signal we simulated a sinusoidal signal with an exponentially decreasing amplitude (case(ii)) and here we observed some sinusoidal variations as shown in Fig 3.2. If we recall the analytic signal method algorithm we have used in vak.c, we calculate IF from the complex analytic signal of the given real signal. An FFT of window size 256 is used to calculate the analytic signal and IF is estimated for all the samples in this window. We called this method a non-sliding window method. To reduce these sinusoidal variations in case(ii) we now modified the procedure in vak.c (to vakman.c) to calculate the IF of only the central sample from the FFT window and shift the window by one sample(to the right) to calculate the IF of the next sample. This modified procedure is termed the sliding window method and it's results can be seen in Fig. 3.3(case(iii)). We increased the FFT window size to 1024 and got a better result, showing that analytic signal method is sensitive to the window size used in calculating the analytic signal. Finally, we applied the algorithm to non-linear variation of frequency as shown in Fig. 3.4(case(iv).

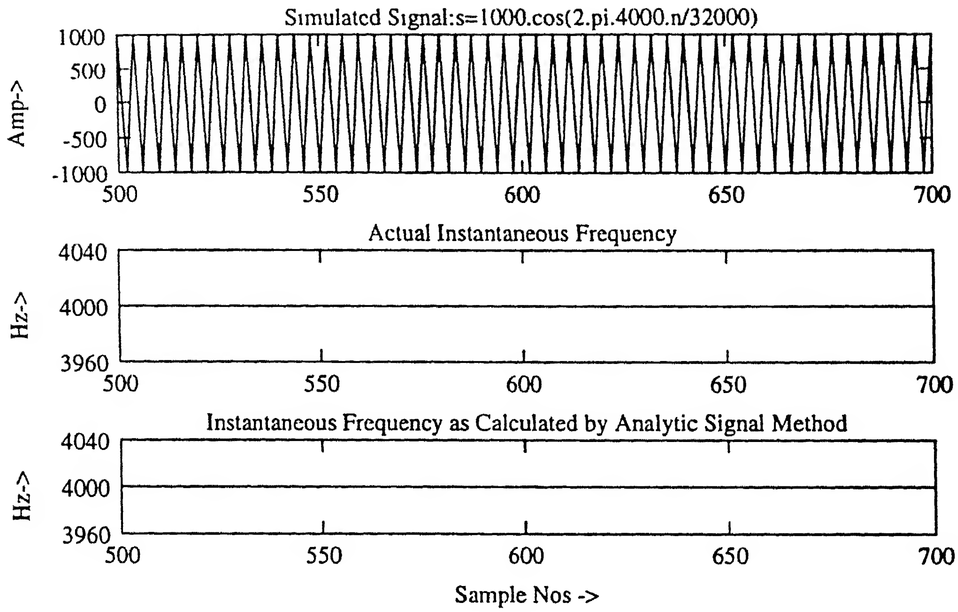

Figure 3.1: case(i) : IF of (constant amplitude, constant frequency) sinusoidal input using 8th order FFT.

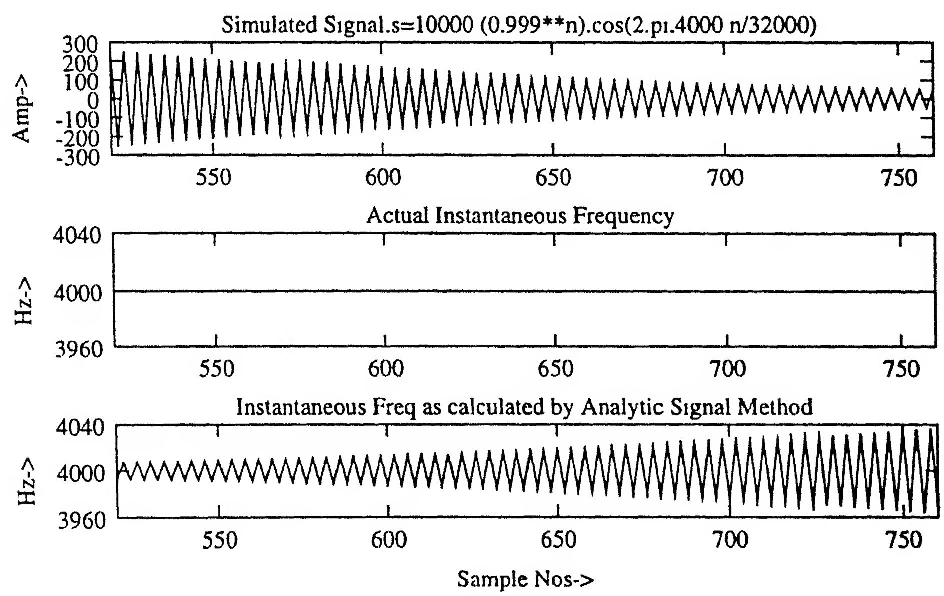

Figure 3.2: case(ii) : IF of (exponentially decreasing amplitude, constant frequency) sinusoidal input using 8th order FFT.

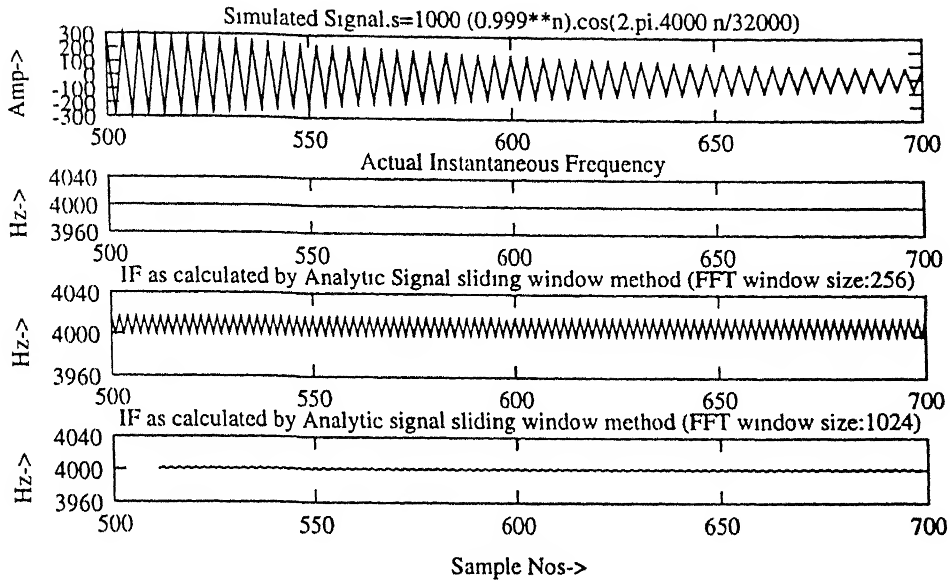

Figure 3.3: case(iii) : IF of (exponentially decreasing amplitude, constant frequency) sinusoidal input using sliding window method.

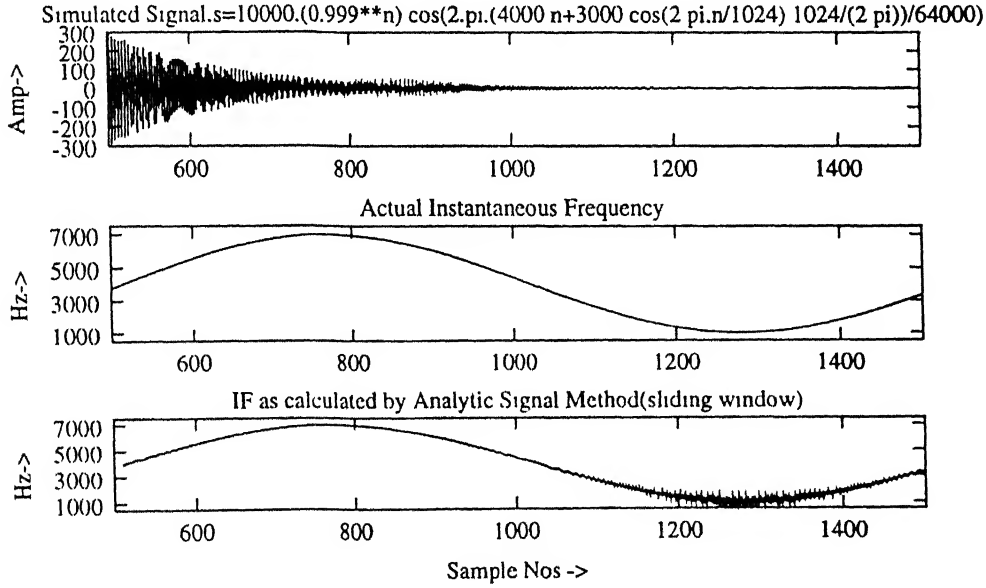

Figure 3.4: case(iv) : IF of (exponentially decreasing amplitude, sinusoidal frequency) sinusoidal input using 10th order FFT sliding window method.

Chapter 4

Application of Analytic Signal Method to Speech Signals

4.1 Method

The analytic signal method of instantaneous frequency estimation is applied on speech resonances taken from the two directories "dr7test" and "dr2test" of the Timit speech database. As we are interested in tracking the first three formant frequencies of each resonance, these three frequencies are calculated for each frame from the LPC and FFT spectra and the AS method algorithm is applied on the band-pass filtered speech at the calculated resonance frequencies. A Gabor band-pass filter is used to band-pass the speech signal and it's impulse response h(n) is given by

$h(n) = exp(-b^2n^2)cos(\Omega_c n)$ where

$b = \alpha T$

$\Omega_c = 2\pi f_c T$

f_c=center frequency of band-pass filter

$\alpha = \sqrt{2\pi}$. RMS bandwidth of Gabor filter

The Gabor band-pass filter is selected to band-pass the speech formants because it avoids producing sidelobes(or big sidelobes after truncation of h(n)) that could produce false pulses in the IF output(further explained in [3]). The RMS filter bandwidth of 400Hz is found to be optimum for widely separated speech formants. The procedure to pre-process the signal is carried out in the following steps.

1. The speech signal is divided into frames of 512 samples each with a frame overlap of 256 samples.

2. The FFT spectrum of order 11 is computed for each frame.

3. The LPC spectrum is computed using order 11 FFT on LPC filter coefficients which are obtained using order 30 LPC analysis on each frame.

4. The first three formant frequencies are calculated for each frame from both the spectra using peak picking.

5. The valid range of speech signal(frames for which the formant frequency is approximately steady in both spectra) and the first three formant frequencies for this range are determined approximately from the values calculated in step 3.

6. The Gabor band-pass filter is applied at the chosen formant frequencies to get the band-pass filtered speech signal at each formant frequency.

The analytic signal method is then applied on the resulting band-pass filtered speech signal to track the formant frequencies.

A command line package is developed which helps us in running the programs(of analytic signal method and other algorithms which help in pre-processing

the signal before applying the AS algorithm)for a given speech signal(described briefly in the appendix).

4.2 Results

Using the procedure outlined we have applied the analytic signal method algorithm first to a sample of synthetic speech and then to speech vowels *ae* and *eh* of both the male and female speakers. The results are given below in the following order.

Sample 1. Synthetic speech sample *ae*.

Sample 2. Speech vowel *ae* of a male speaker.

Sample 3. Speech vowel *ae* of a female speaker.

Sample 4. Speech vowel *eh* of male speaker.

Sample 5. Speech vowel *eh* of a female speaker.

Sample 6. Speech vowel *ae* of a male speaker.

Sample 1.

Formants from LPC and FFT spectra

Formant 1 : 660 Hz

Formant 2 : 1500 Hz

Formant 3 : 2400 Hz

The results can be seen in Fig.5.1

Sample 2.

`sample taken from c:\speech\data\dr7test\mdvc0\sa2(49260-52440)`

Calculated formants from LPC and FFT spectra

Formant 1 : 710 Hz

Formant 2 : 1770 Hz

Formant 3 : 2470 Hz

Considered range of speech samples : 700 - 1500 sample nos

The results can be seen in Fig.5.2

Sample 3.

`Sample taken from c:\speech\data\dr7test\fdhc0\sa2(27266-29639)`

Calculated formants from LPC and FFT spectra

Formant 1 : 800 Hz

Formant 2 : 1800 Hz

Formant 3 : 2600 Hz

Considered range of speech samples : 650 - 1250 sample nos

The results can be seen in Figure 5.3

Sample 4

Sample taken from c:\speech\data\dr7test\mdvc0\si2196(8930-10750)

Calculated formants from LPC and FFT spectra

Formant 1 : 610 Hz

Formant 2 : 1530 Hz

Formant 3 : 2350 Hz

Considered range of speech samples : 400 - 1050 sample nos

The results can be seen in Fig.5.4

Sample 5.

Sample taken from c:\speech\data\dr7test\fcau0\si1667(42670-43934)

Calculated formants from LPC and FFT spectra

Formant 1 : 718 Hz

Formant 2 : 1660 Hz

Formant 3 : 2800 Hz

Considered range of speech samples : 450 - 850 sample nos

The results can be seen in Fig.5.5

Sample 6.

```
Sample taken from  c:\speech\data\dr2test\mpdf0\sa2(5910-8680)
```

Calculated formants from LPC and FFT spectra

Formant 1 : 625 Hz

Formant 2 : 1620 Hz

Formant 3 : 2370 Hz

Considered range of speech samples : 450 - 850 sample nos

The results can be seen in Fig.5.6

4.3 Observations and Conclusions

We can observe large pulses at periods of pitch in all the plots which can be attributed to the excitation at the glottal closure instants. The estimate of instantaneous frequency also is perturbed in the vicinity of glottal opening (in sample 2). Hence instantaneous frequency estimates are not reliable in regions close to the glottal closure and glottal opening instants. Comparing the plots with the approximate glottal closure-opening instant waveform we can see that (in samples 2 and 4) the frequency remains constant in glottal closure period and increases in the glottal opening time. This is in line with the observation in [6] that formant frequency and bandwidth are higher in the glottal opening time in a pitch period than in the glottal closure time. The small sinusoidal variations present in some of the plots (sample 1 and 3) may be due to the limited duration of the window size (1024 sample nos.) taken in calculating the complex analytic signal of the given real signal.

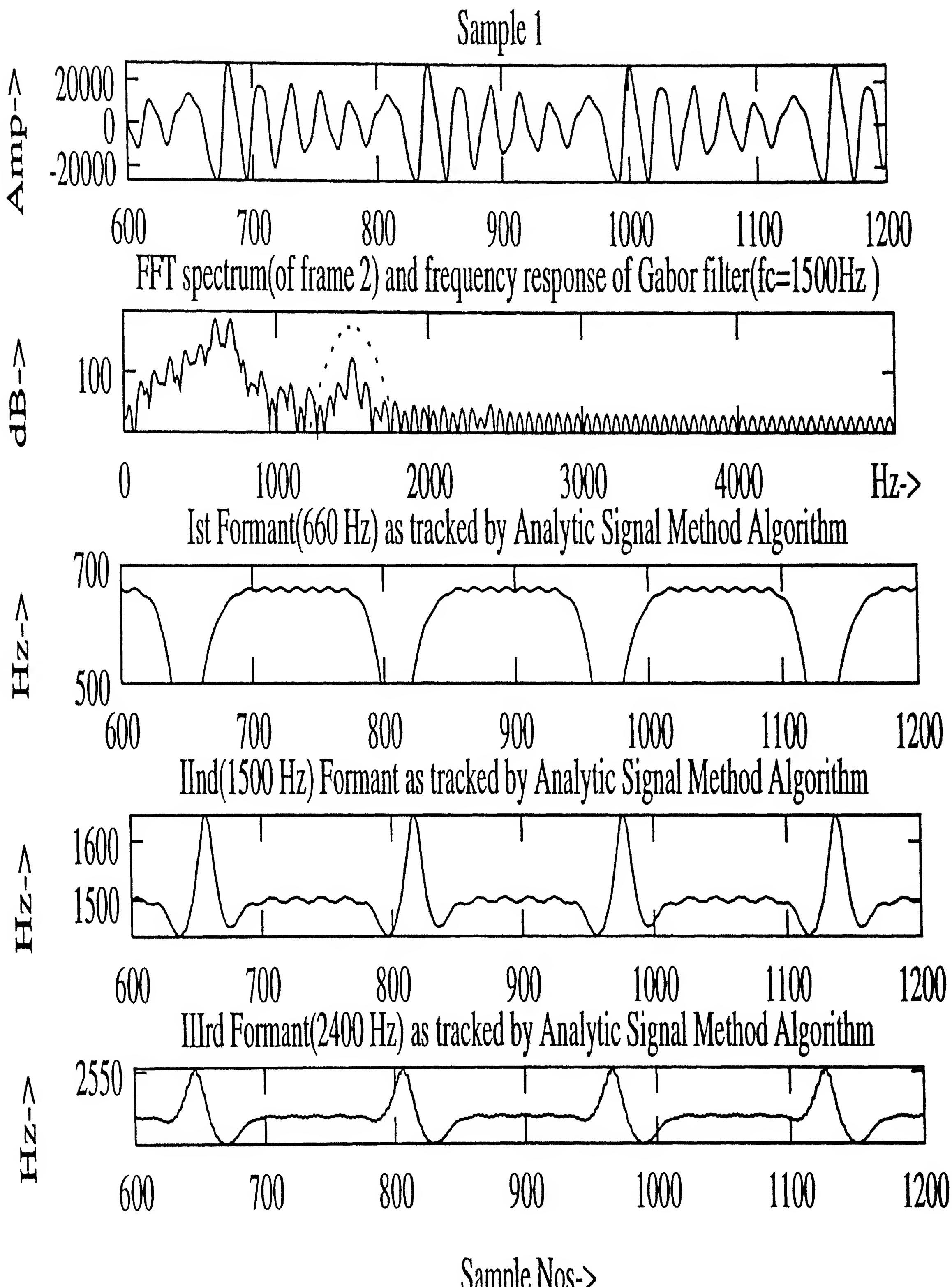

Figure 4.1: IF output of Sample 1

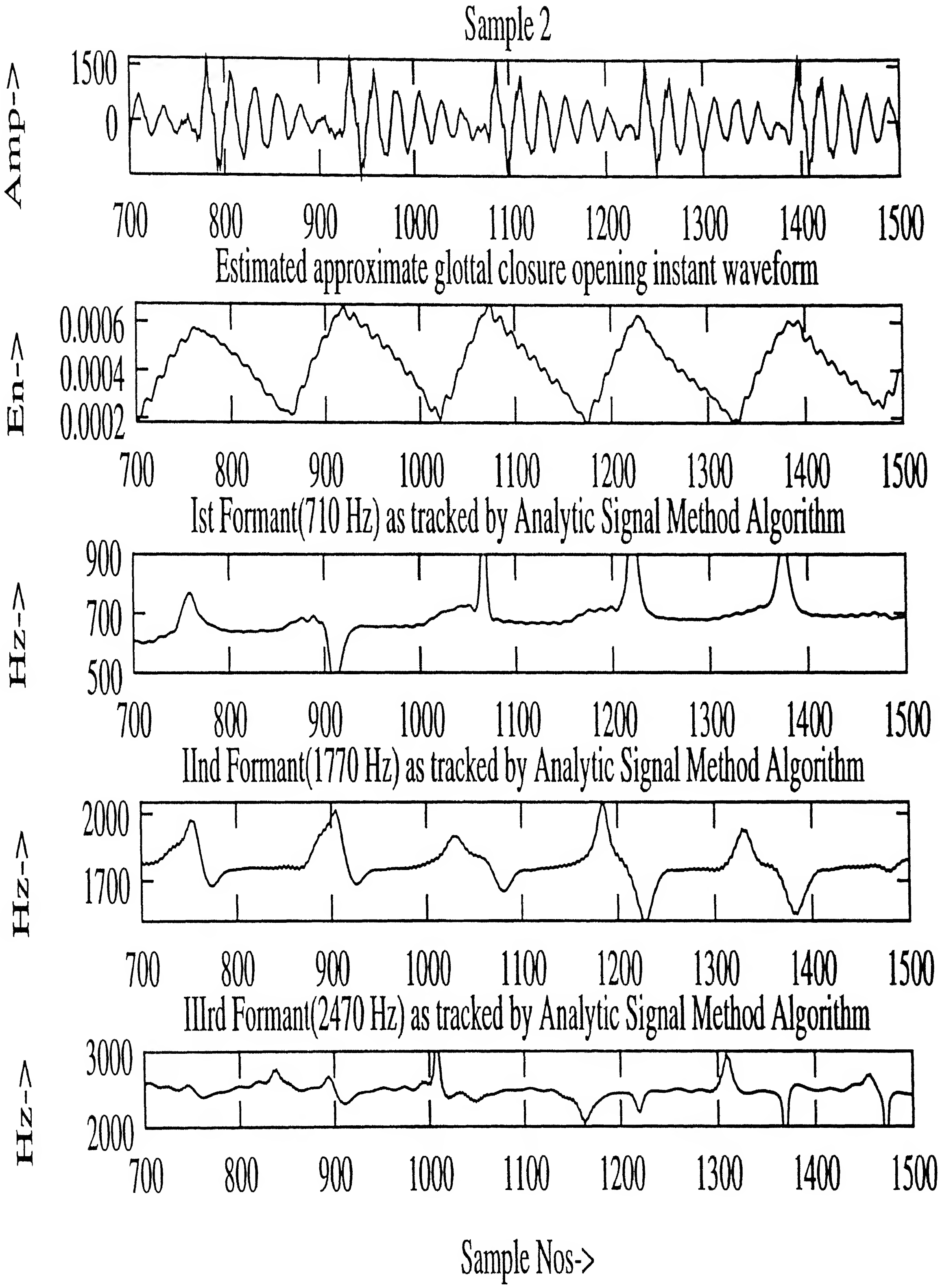

Figure 4.2: IF output of Sample 2

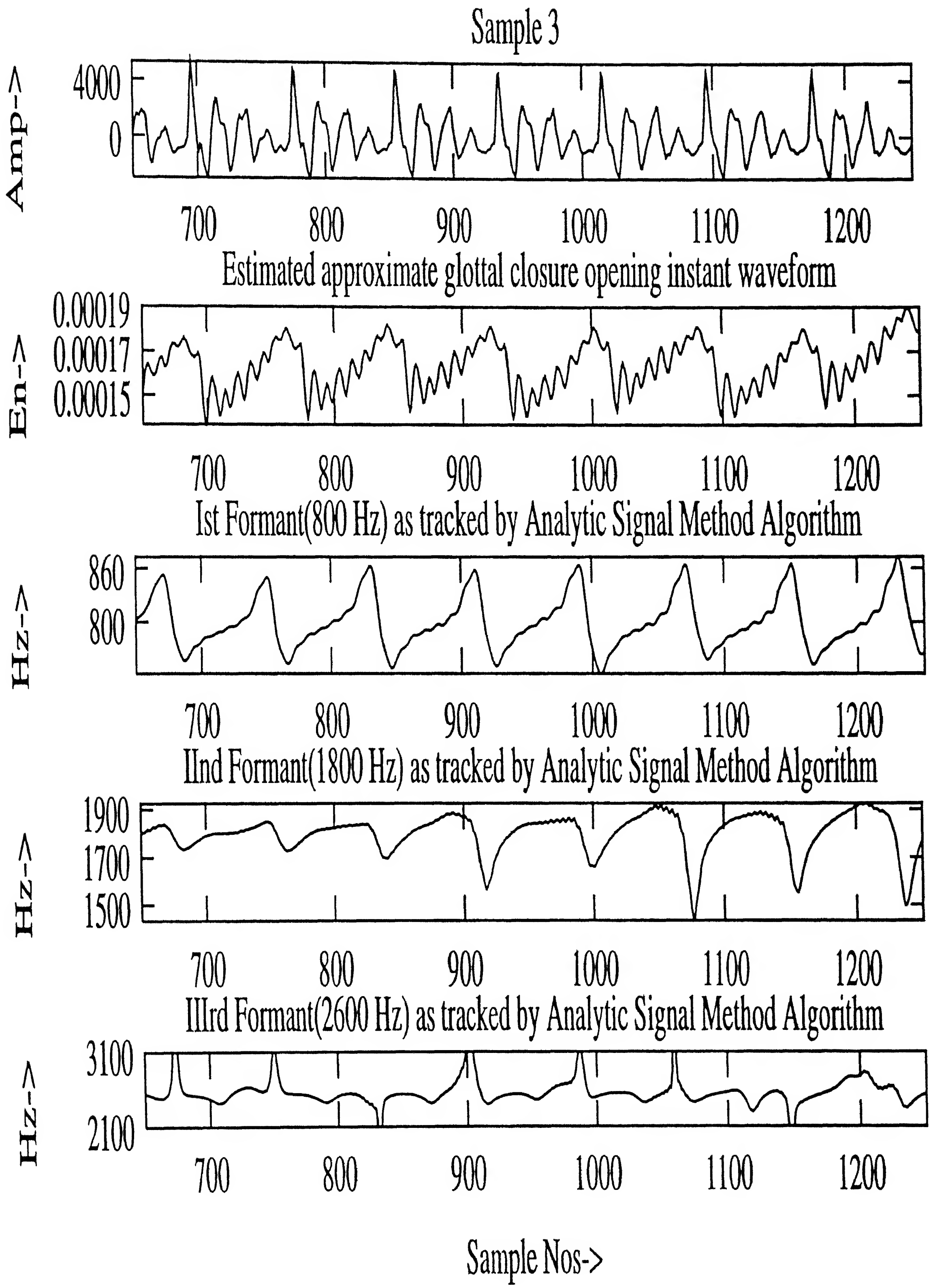

Figure 4.3: IF output of Sample 3

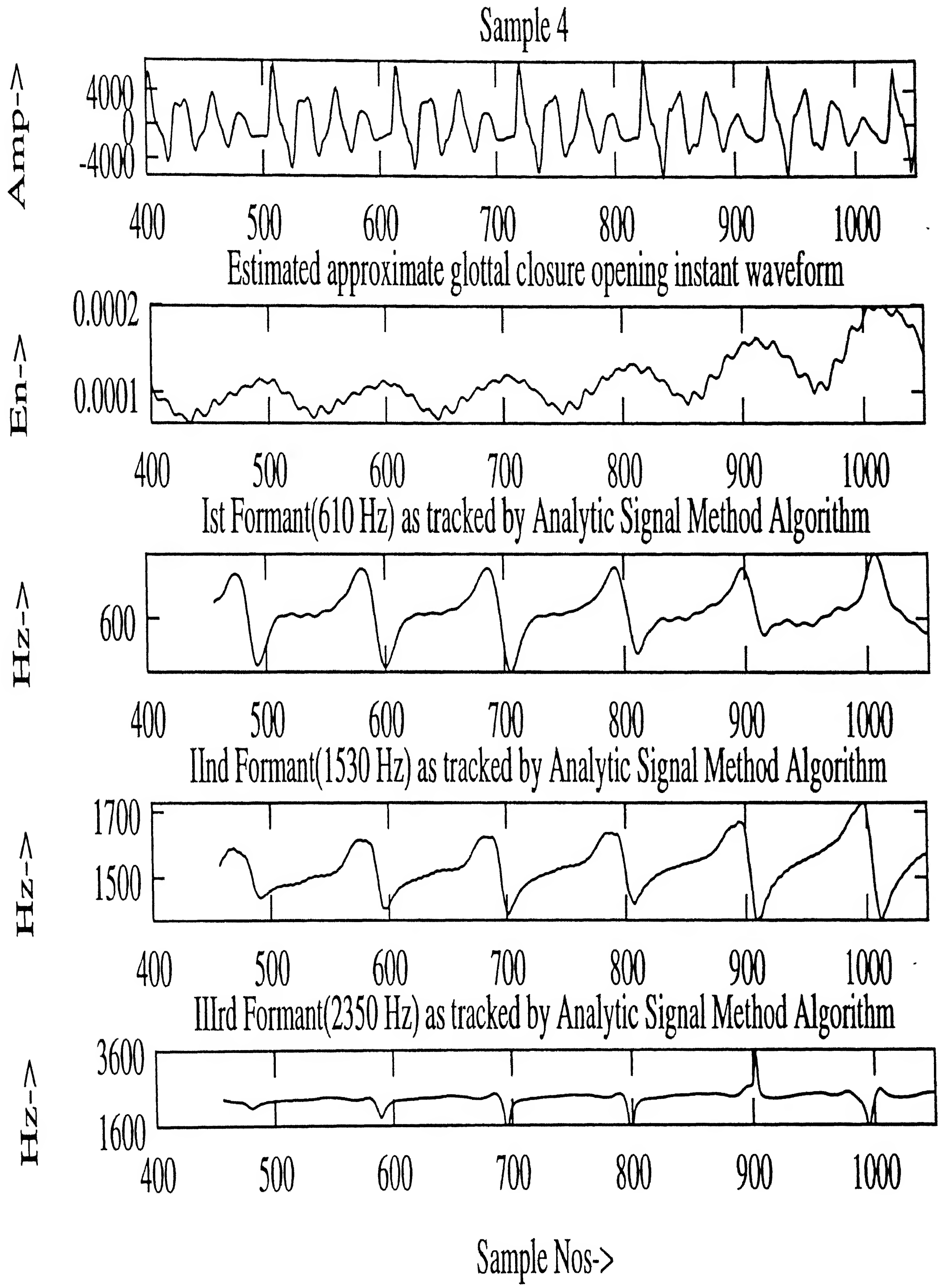

Figure 4.4: IF output of Sample 4

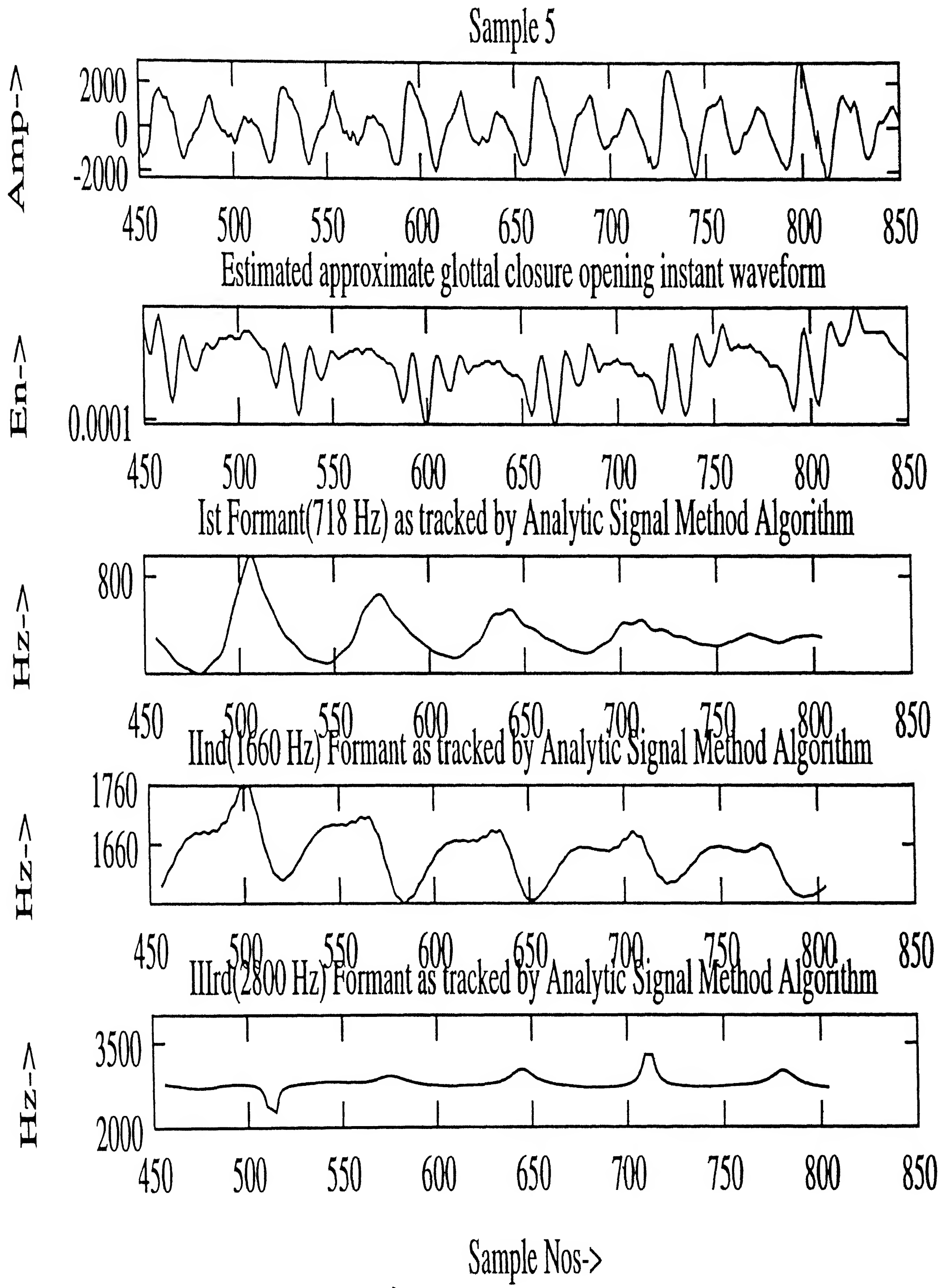

Figure 4.5: IF output of Sample 5

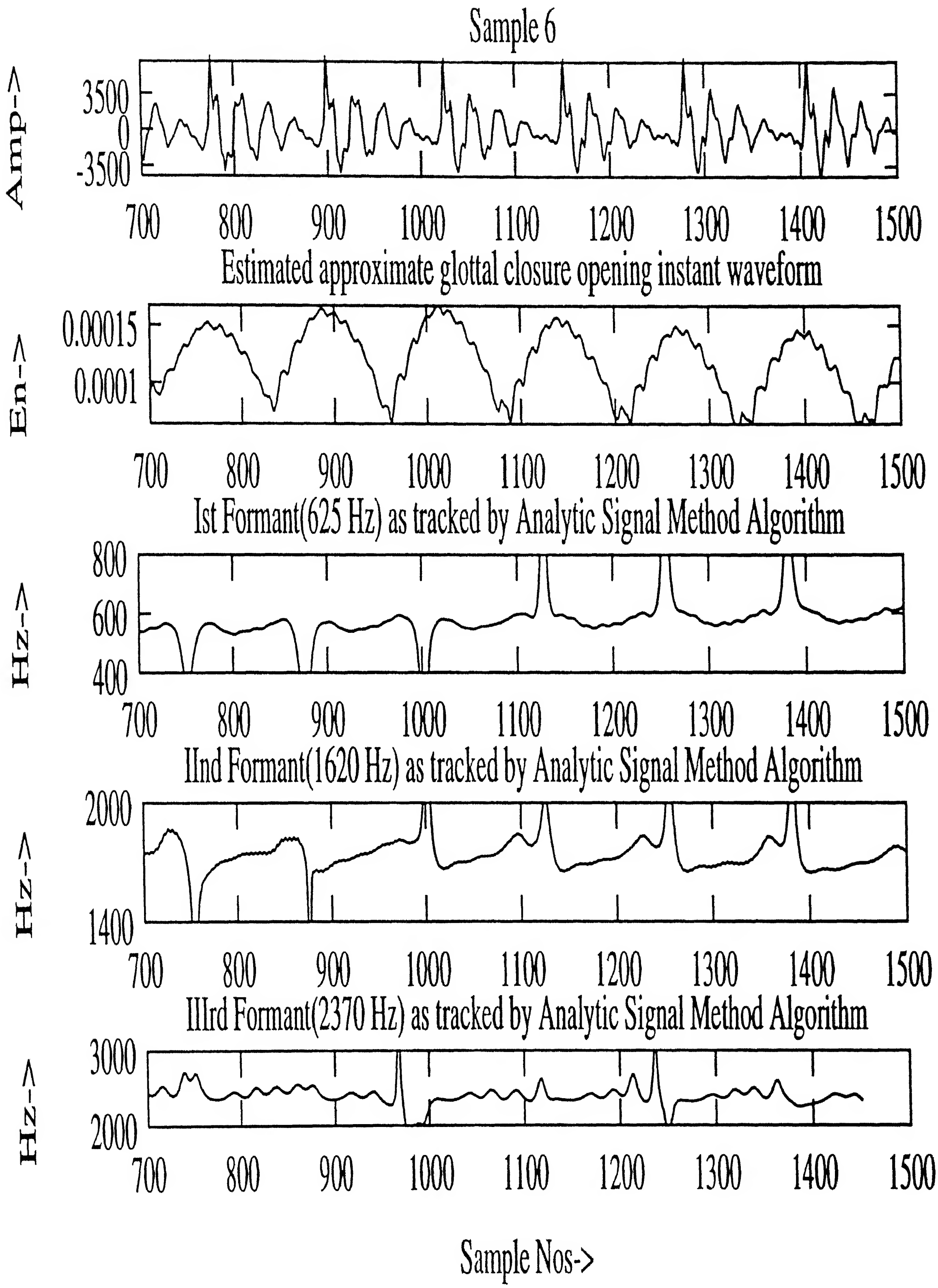

Figure 4.6. IF output of Sample 6

Appendix : Description of the Package Developed to Apply AS Algorithm to Speech Signal

The following is the list of main C programs written to pre-process the speech signal and apply the analytic signal algorithm on it.

1. vak.c (program implementing the analytic signal algorithm with out sliding the window)

2. vakman.c (program implementing the analytic signal algorithm by sliding the window)

3. filter.c (program to calculate the output signal which is obtained after convolving the input signal with the impulse response of Gabor band-pass filter of given center frequency and band-width)

4. fftspec.c (This program calculates the 11th order FFT spectrum of the input signal)

5. lpcspec.c and analysis.c (The program lpcspec.c calculates the 11th order FFT spectrum of the LPC filter co-efficients which are obtained after analysing the input signal using analysis.c)

6. eagcoi.c (This program uses the algorithm presented in [5] to calculate the approximate glottal closure opening instants)

The above programs (along with some small programs like those which find out the number of samples in a particular speech file) are compiled on a HP-UX Workstation using ANSI C compiler and are run with the help of following shell programs (gnuplot is used to view the plots)

1. vsg (Shell program used to view the input speech signal(tim file) in gnuplot)

2. clfs (Program which calculates the LPC and FFT spectra using the above mentioned C programs analysis.c, lpcspec.c and fftspec.c)

3. vspcaf (Program to view the calculated FFT and LPC specta and compute the 3 formants for each frame using peak-picking algorithm)

4. fpv (Program to calculate the IF output after filtering by applying the analytic signal algorithm and storing the calculated output in the output directory)

5. mkd (Program to create a special directory for each sample and store the sample in it. It also creates an output directory for each sample in its directory).

Example showing the procedure to run the above package

Let us assume that all the executable programs and the shell programs are stored in directory "/usr/as" and the speech signal ss.tim(tim file taken from Timit data-base) is also in this directory. The following is the list of commands which are to be executed and a brief description as to what happens after their execution

/usr/as≫ mkd ss

This command creates a directory "ss" in the present directory and stores the speech

sample ss.tim in it. It also creates an output directory "/usr/as/ss/output" to store the outputs of this sample in this directory.

/usr/as/as≫ vsg ss

This command opens a gnuplot window on screen and display's the signal in it.

/usr/as/as≫ clfs ss

This command calculates the LPC and FFT spectra of the speech signal and stores them in the present directory.

/usr/as/as≫ vspcaf ss 10

We will be knowing the no of frames of the speech signal from the previous command "clfs ss" and no of frames is given as a parameter in the above command to see the LPC and FFT spectra of each frame in the gnu-plot while pressing ENTER key simultaneously after viewing each plot. This shell program also simultaneously calculates the formant frequencies for each frame and display's them on the screen and also stores them in a file for later reference.

/usr/as/as≫ fpv ss fc(bpf center frequency fc)

This program runs the filter program for given center frequency and applies the AS method on the resulting signal. It also moves the output file to the output directory.

Bibliography

[1] D.Vakman," On the Analytic Signal, the Teager-Kaiser Energy Algorithm, and Other Methods for Defining Amplitude and Frequency",*IEEE Trans. on Signal Processing*,vol. 44,no. 4,pp 791-797,April 1996.

[2] D.G.Childers and C.F.Wong,"Measuring and Modelling Vocal Source-Tract Interaction",*IEEE Trans. on Biomedical Engineering*,vol. 41,no. 7,pp 663-671 , July 1994.

[3] P.Margos, J.F.Kaiser and T.F.Quatieri,"Energy Separation in Signal Modulations with Application to Speech Analysis",*IEEE Trans. on Signal Processing*, vol. 41,no. 10,pp 3024-3051,October 1993.

[4] P.Margos, A.Potamianos,"A Comparison of the Energy Operator and the Hilbert Transform approach to Signal and Speech Demodulation",*Signal Processing*,pp 95-120,1994.

[5] S.Parathasarathy and D.W.Tufts,"Excitation-Synchronous Modeling of Voiced Speech",*IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. ASSP-35,no. 9,September 1987.

[6] B.Yegnarayana and P.S.Murthy,"Source-System Windowing for Speech Analysis and Synthesis",*IEEE Trans. on Speech and Audio Processing*,vol. 4,no. 2, March 1996.

[7] B.Boashash,"Estimating and Interpreting The Instantaneous Frequency of a Signal- Part I:Fundamentals", *Proc. of IEEE*,vol. 80,no. 4,April 1992.

[8] B.Boashash,"Estimating and Interpreting The Instantaneous Frequency of a Signal- Part II:Fundamentals", *Proc. of IEEE*,vol. 80,no. 4,April 1992.

[9] L.R.Rabiner and R.W.Schafer,*Digital Processing of Speech Signals*, Englewood Cliffs, NJ:Prentice-Hall,1978.

[10] R.W.Schafer and J.D.Markel,*Speech Analysis*,New York:IEEE Press,1979.

[11] Roman Kuc,*Introduction to Digital Signal Processing*,McGraw-Hill Book Company, 1988.

[12] Y.Kanetkar,*Let Us C*,BPB Publications,1995.

[13] B.V.Kernighan and R.Pike,*The Unix Programming Environment*,New Delhi:PHI Pvt. Ltd.,1996.

[14] S.G.Kochan and P.H.Wood,*Exploring the Unix System*,Delhi:CBS Pubs. and Distrbs. ,1987.

[15] G.Fant and Q.G.Lin,"Glottal Source-Vocal Tract Acoustic Interaction", STL-QPSR 1/1987,Royal Institute of Technology,Stockholm,Sweden,pp 13-27, 1987.

[16] H.M.Teager,"Physiology of Speech Phoneme Production",3rd Tech. Rep., Contract MDA 904-81-C-0413,1981.